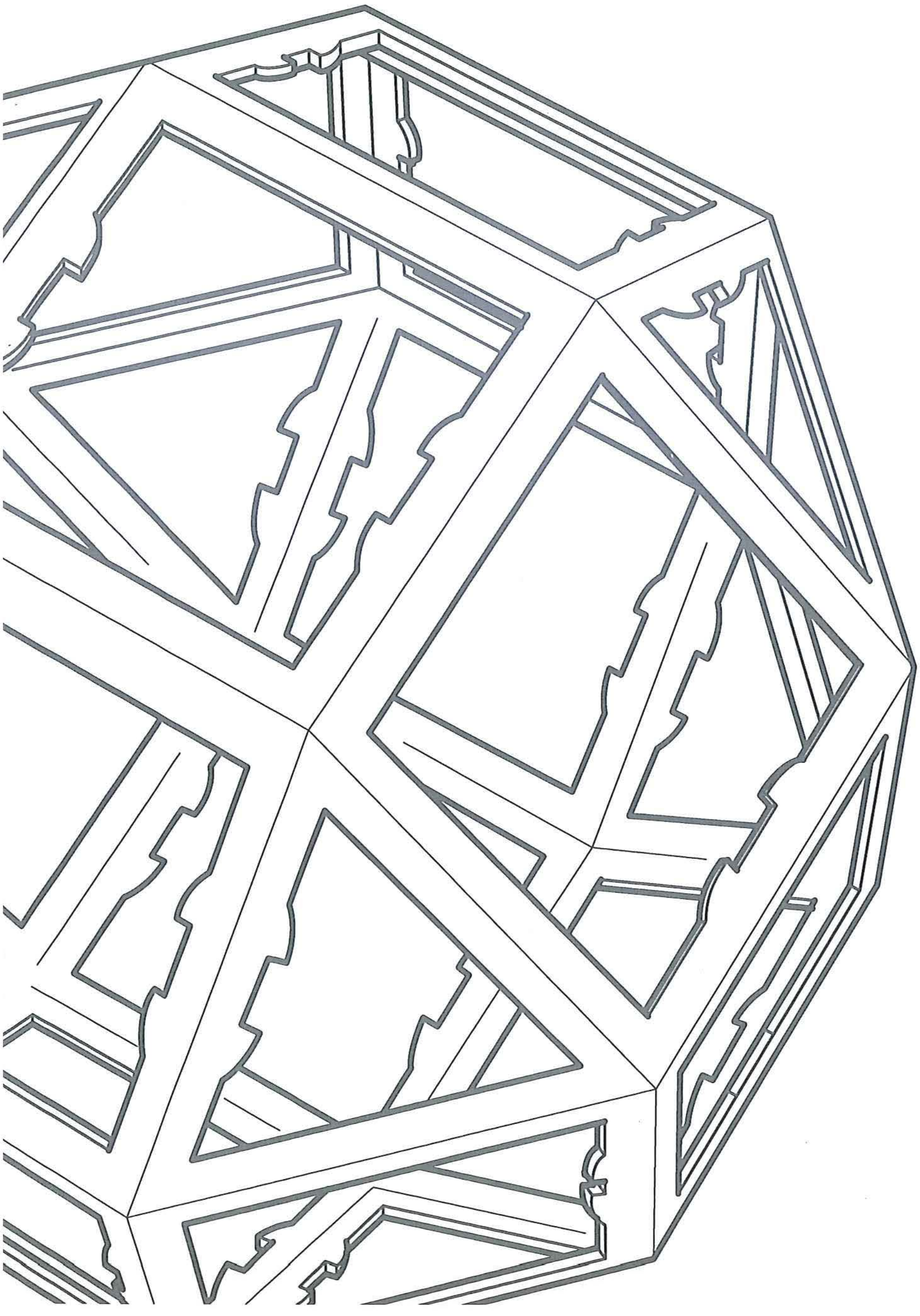
Department of Mathematics and Statistics
College of Engineering

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Summer Research Project

A Brief Interaction with Continued Fractions

By William Frost



09

University of Canterbury

Math305 Summer Project

A Brief Interaction with Continued Fractions

Author:
William Frost

Supervisor:
Dr. Rua Murray

Abstract

Basic concepts of simple continued fractions are introduced and some important theorems explored. The effect of an infinite continued fraction's elements forming a convergent series is looked at via an example of geometric series. The Gauss-Kuzmin-Wirsing operator, operating on functions on the interval $[0,1]$, is studied numerically. Its invariant density is explored and the rate at which an initial density transforms into the invariant density shown to be $O(e^{-Sn})$ for iteration n. The transformation associated with the operator is applied numerically to a single random point in $[0,1]$ and interpretations of the results given.

February 5, 2009

1 Introduction

The history of continued fractions is very long and rich, beginning with "Euclid's algorithm for the greatest common divisor at least three centuries B.C" [3]. Many great mathematicians, including Gauss have played a part in the construction of the knowledge of continued fractions we have today. I will be looking at a small portion of this expanse of knowledge.

This report considers simple continued fractions. In section 2, an explanation of some background concepts of the simple continued fraction is given, introducing the reader to mathematical terms such as *remainders*, *sections* and *convergents*; examples of each are given. In section 3, elements of an infinite continued fraction whose series is convergent are explored, showing the effects graphically and giving a brief explanation. Finally, section 4 looks at the Gauss-Kuzmin-Wirsing operator, its effect on the density of numbers in the interval [0,1] and effect of its corresponding transformation on random single points in the interval [0,1]. All except a minority of initial points in [0,1], are shown to give the same final time averaged probability density. This section also looks at the rate at which an initial probability density is transformed into an invariant density by the Gauss-Kuzmin-Wirsing operator. The main resource used for my understanding of continued fractions was A. I. Khinchin's book "Continued Fractions" [7].

2 Basics of Continued Fractions

2.1 Method of finding the Continued Fraction expansion of a number

Any real number x can be represented efficiently by a simple continued fraction expansion:

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}}$$

To find the continued fraction expansion of x we use the following quite natural procedure:

Let us denote the largest integer not exceeding x by a_0 ($a_0 = \lfloor x \rfloor$).
We can write:

$$x = a_0 + \frac{1}{r_1}\ , \tag{1}$$

where $\frac{1}{r_1}$ is the fractional part of x. Thus $\frac{1}{r_1} = x - a_0$. We now repeat this process with r_1 in place of x.
Denote by a_1 the largest integer not exceeding r_1, so

$$r_1 = a_1 + \frac{1}{r_2}\ .$$

In general we have

$$r_n = a_n + \frac{1}{r_{n+1}}\ .$$

Thus x can be written

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{}{\ddots \, \frac{1}{r_{n+1}}}}} .$$

The symbols $a_0, a_1, a_2, \ldots$ are called the *elements* of the continued fraction. The process above is repeated until either r_{n+1} is found to be an integer, thus $a_{n+1} = r_{n+1}$ and the iteration stops, or it may continue indefinitely, resulting in an infinite number of elements.

The result of this process is either

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{}{\ddots \, \frac{1}{a_{n+1}}}}} ,$$

a terminating continued fraction, or

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{}{\ddots}}} ,$$

an infinite continued fraction (respectively).

Khinchin's book [7] contains a theorem explaining the difference between these two types of continued fractions.

Theorem 1 ([7, Theorem 14]). *To every real number α there corresponds, uniquely, a continued fraction whose value is this number. If α is rational this continued fraction terminates. If α is irrational, it has an infinite expansion.*

Note: uniqueness is a consequence of the simple continued fraction expansion. The process used to find this expansion cannot possibly give multiple results for the elements a_i, as the largest integer not exceeding r_i is unique (for $i = 1, 2, \ldots$).

Examples

Rational: $x_{rat} = \frac{7}{5}$

$$\begin{aligned} x_{rat} &= \frac{7}{5} = 1 + \frac{2}{5} \\ &= 1 + \frac{1}{\frac{5}{2}} = 1 + \frac{1}{2 + \frac{1}{2}} \end{aligned}$$

Irrational: $x_{irr} = \sqrt{2}$

$$\begin{aligned} x_{irr} &= \sqrt{2} = 1 + (\sqrt{2} - 1) \\ &= 1 + \frac{1}{\frac{1}{\sqrt{2}-1}} = 1 + \frac{1}{1 + \sqrt{2}} \\ &= 1 + \frac{1}{2 + (\sqrt{2} - 1)} \end{aligned}$$

This process repeatedly gives an output of $(\sqrt{2}-1)$ for the fractional part remaining at each step. It logically follows that we end up with a repeated output of 2 for each element a_i, $i = 1, 2, \ldots$. The process repeats indefinitely as we will always get a fractional part remaining of $(\sqrt{2}-1)$. Thus, $\sqrt{2}$ has the infinite continued fraction expansion:

$$\sqrt{2} = 1 + \cfrac{1}{2 + \frac{1}{2 + \frac{1}{2+\ldots}}}$$

2.2 Simple Continued Fractions

The form of the continued fractions dealt with in this report are known as "simple" [7]. These continued fractions are, as above, of the form:

$$\alpha = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2+\ldots}} \ , \tag{2}$$

which can be represented by the simpler notation

$$\alpha = [a_0; a_1, a_2, a_3, \ldots] \ . \tag{3}$$

In this report, the elements a_1, a_2, etc. will be *natural* numbers, while the element a_0, can take any integer value.

Definitions

Some important mathematical quantities when dealing with continued fractions follow:

Remainder. The k^{th} remainder is defined as:

$$r_k = [a_k; a_{k+1}, \ldots, a_n] \text{ where } 0 < k \leq n \tag{4}$$

(for terminating continued fractions)

$$r_k = [a_k; a_{k+1}, \ldots] \tag{5}$$

(for infinite continued fractions)

Section. The k^{th} section is defined as:

$$s_k = [a_0; a_1, a_2, \ldots, a_k] \tag{6}$$

(for both infinite and terminating continued fractions)

These terms will be used throughout this report.

Example

A nice example of a continued fraction is the golden ratio, $\Phi = \frac{\sqrt{5}+1}{2}$ as it has the simple continued fraction expansion $\Phi = [1; 1, 1, 1, 1, 1, \ldots]$. The argument is similar to the $\sqrt{2}$ example in subsection 2.1 and exploits the fact that $\Phi = 1 + \frac{1}{\Phi}$. It is an irrational number, so the ones are repeated indefinitely. Because of this repetition of ones Φ is known as having a *periodic continued fraction expansion.*

The n^{th} remainder of Φ for any n is $\Phi = [1; 1, 1, 1, 1, 1, \ldots]$.
The n^{th} section of Φ is $S_n = [1; 1, 1, 1, \ldots, 1]$ ($n+1$ 1s, a_0 to a_n).

What value does the n^{th} section have?

Table 1

n	S_n	value of S_n	$\lvert S_n - \Phi\rvert$ (4 s.f.)
0	$[1]$	1	6.180×10^{-1}
1	$[1; 1]$	2	3.820×10^{-1}
2	$[1; 1, 1]$	$\frac{3}{2}$	1.180×10^{-1}
3	$[1; 1, 1, 1]$	$\frac{5}{2}$	4.863×10^{-2}
4	$[1; 1, 1, 1, 1]$	$\frac{8}{5}$	1.8803×10^{-2}
5	$[1; 1, 1, 1, 1, 1]$	$\frac{13}{8}$	6.966×10^{-3}
6	$[1; 1, 1, 1, 1, 1, 1]$	$\frac{21}{13}$	2.649×10^{-3}
7	$[1; 1, 1, 1, 1, 1, 1, 1]$	$\frac{34}{21}$	2.649×10^{-3}
8	$[1; 1, 1, 1, 1, 1, 1, 1, 1]$	$\frac{55}{34}$	1.014×10^{-3}
9	$[1; 1, 1, 1, 1, 1, 1, 1, 1, 1]$	$\frac{89}{55}$	3.869×10^{-4}
10	$[1; 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$	$\frac{144}{89}$	5.646×10^{-5}

Table 1 shows the first 7 sections and also S_{10} of Φ along with their values and difference from Φ, if we were to continue this list we would find that $\lim_{n\to\infty} |S_n - \Phi| = 0$, thus $\lim_{n\to\infty} S_n = \Phi$, and $\Phi = 1.6180339\ldots$. Note that the error in the last column $|S_n - \Phi|$, is very small in comparison with the reciprocal of the denominator in S_n. In general if you approximate an irrational x by a rational $\frac{p}{q}$ you might expect $\left|x - \frac{p}{q}\right| \approx \frac{1}{q}$, but S_n seems much better.

The nature of the values of the sections is looked at in the next subsection.

2.3 Convergents

For finite k, the k^{th} section can be written as the fraction $\frac{p_k}{q_k} = [a_0; a_1, \ldots, a_k]$. Because the elements are all integers, $\frac{p_k}{q_k}$ is rational and so both p_k and q_k can be chosen as integers dependent on $a_0, a_1, a_2, \ldots, a_k$. $\frac{p_k}{q_k}$ is known as the k^{th} *Convergent.*

The following theorem and corollaries involve the convergents and can give a good understanding of the workings of continued fractions.

Theorem 2 ([7, Theorem 1]). *For any $k \geq 1$, if we define*

$$p_k = a_k p_{k-1} + p_{k-2} \tag{7}$$
$$q_k = a_k q_{k-1} + q_{k-2} \tag{8}$$

(with $p_{-1} = 1$, $q_{-1} = 0$, $p_0 = a_0$, $q_0 = 1$) then we have

$$\frac{p_k}{q_k} = [a_0; a_1, a_2, \ldots, a_k].$$

Proof. (by induction, adapted from [10])

We prove by induction that for any X, $[a_0; a_1, \ldots, a_{k-1}, X] = \frac{Xp_{k-1}+p_{k-2}}{Xq_{k-1}+q_{k-2}}$

Case $k = 1$:

$$\begin{aligned} [a_0, X] &= a_0 + \frac{1}{X} && (9) \\ &= \frac{Xa_0 + 1}{X \cdot 1 + 0} && (10) \\ &= \frac{Xp_0 + p_{-1}}{Xq_0 + q_{-1}} && (11) \end{aligned}$$

Thus true for $k = 1$.

The k^{th} case, for the induction step, assume $[a_0; a_1, \ldots, a_{k-1}, X] = \frac{Xp_{k-1}+p_{k-2}}{Xq_{k-1}+q_{k-2}}$ is true.

So, to prove $(k+1)^{th}$, we use $X' = a_k + \frac{1}{X}$.

$$\begin{aligned} [a_0; a_1, \ldots, a_{k-1}, a_k, X] &= [a_0; a_1, a_2, a_3, \ldots, a_k + \frac{1}{X}] \\ &= [a_0; a_1, a_2, a_3, \ldots, a_{k-1}, X'] \\ &= \frac{X'p_{k-1} + p_{k-2}}{X'q_{k-1} + q_{k-2}} \\ &= \frac{(a_k + \frac{1}{X})p_{k-1} + p_{k-2}}{(a_k + \frac{1}{X})q_{k-1} + q_{k-2}} \\ &= \frac{(a_kX + 1)p_{k-1} + Xp_{k-2}}{(a_kX + 1)q_{k-1} + Xq_{k-2}} \\ &= \frac{X(a_kp_{k-1} + p_{k-2}) + p_{k-1}}{X(a_kq_{k-1} + q_{k-2}) + q_{k-1}} \\ &= \frac{Xp_k + p_{k-1}}{Xq_k + q_{k-1}} \,. \end{aligned}$$

The theorem follows. □

[We note that the convergent $\frac{p_{-1}}{q_{-1}}$ has no real meaning, it just allows the theorem to work for $k \geq 1$.]

The following corollaries are written in terms of Theorem 2's notation.

Corollary 3. *For all $k \geq 0$*

$$q_kp_{k-1} - p_kq_{k-1} = (-1)^k \qquad (12)$$

Proof. Case $k = 0$,

$$q_0 p_{-1} - p_0 q_{-1} = 1 = (-1)^0.$$

Thus true for case $k = 0$. Given equations (7) and (8),

$$\begin{aligned} q_k p_{k-1} - p_k q_{k-1} &= p_{k-1}(a_k q_{k-1} + q_{k-2}) - q_{k-1}(a_k p_{k-1} + p_{k-2}) \\ &= a_k p_{k-1} q_{k-1} + p_{k-1} q_{k-2} - a_k p_{k-1} q_{k-1} - q_{k-1} p_{k-2} \\ &= -(q_{k-1} p_{k-2} - p_{k-1} q_{k-2}) \end{aligned}$$

This shows that consecutive values of $q_k p_{k-1} - p_k q_{k-1}$ differ by a factor of -1 and the first case gives 1, so the corollary is proved. □

Corollary 4. *For all $k \geq 0$*

$$\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} = \frac{(-1)^k}{q_k q_{k-1}} \tag{13}$$

Proof. Divide equation (12) by $q_k q_{k-1}$. □

Corollary 5. *For all $k \geq 1$*

$$q_k p_{k-2} - p_k q_{k-2} = (-1)^{k-1} a_k \tag{14}$$

Proof. Case $k = 1$:

$$q_1 p_{-1} - p_1 q_{-1} = a_1 = a_1(-1)^0.$$

Thus true for case $k = 1$. Given equations (7) and (8),

$$\begin{aligned} q_k p_{k-2} - p_k q_{k-2} &= p_{k-2}(a_k q_{k-1} + q_{k-2}) - q_{k-2}(a_k p_{k-1} + p_{k-2}) \\ &= a_k p_{k-2} q_{k-1} + p_{k-2} q_{k-2} - a_k p_{k-1} q_{k-2} - q_{k-2} p_{k-2} \\ &= a_k(q_{k-1} p_{k-2} - p_{k-1} q_{k-2}) \ . \end{aligned}$$

In Corollary 3 we have shown that $q_k p_{k-1} - p_k q_{k-1} = (-1)^k$, so from this final statement $q_{k-1} p_{k-2} - p_{k-1} q_{k-2} = (-1)^{k-1}$.

Therefore,

$$q_k p_{k-2} - p_k q_{k-2} = (-1)^{k-1} a_k,$$

proving the corollary. □

Corollary 6. *For all $k \geq 1$*

$$\frac{p_{k-2}}{q_{k-2}} - \frac{p_k}{q_k} = \frac{(-1)^{k-1} a_k}{q_k q_{k-2}} \tag{15}$$

Proof. Divide equation (14) by $q_k q_{k-2}$. □

We can understand a lot about the convergents using the above corollaries.

First of all, Corollary 3 tells us that p_k and q_k must be co-prime and so the convergents are in reduced form. To show this, let us assume that they are not co-prime. So we can write $p_k = as$ and $q_k = at$, for some common factor a, where a, s and t are all integers. Following Corollary 3, equation (12) becomes $p_{k-1}at - q_{k-1}as = (-1)^k$, thus $a(p_{k-1}t - q_{k-1}s) = (-1)^k$. But this means that a must be a factor of $(-1)^k$, thus $a = \pm 1$. This is a contradiction, so p_k and q_k are co-prime and thus $\frac{p_k}{q_k}$ is in reduced form.

By combining Corollaries 4 and 6, we can get a nice picture of what happens to the sequence of convergents from the 0^{th} convergent, $\frac{p_0}{q_0}$, through to the k^{th} convergent, $\frac{p_k}{q_k}$. We find that we can partition the types of convergents into two categories, "odd" and "even," due to similar characteristics. Odd convergents involve convergents with odd values of k and even convergents, even values of k respectively.

From Corollary 6, odd k gives the right hand side (RHS) of the equation a positive value and thus

$$\begin{aligned} \frac{p_{k-2}}{q_{k-2}} - \frac{p_k}{q_k} &> 0 \\ \frac{p_{k-2}}{q_{k-2}} &> \frac{p_k}{q_k}, \end{aligned}$$

showing that consecutive **odd** convergents are smaller, so the odd convergents form a decreasing sequence. Whereas an even k gives the RHS a negative value and thus

$$\begin{aligned} \frac{p_{k-2}}{q_{k-2}} - \frac{p_k}{q_k} &< 0 \\ \frac{p_{k-2}}{q_{k-2}} &< \frac{p_k}{q_k}, \end{aligned}$$

showing that consecutive **even** convergents are larger, so the even convergents form an increasing sequence. From Corollary 4, odd k gives the RHS of the equation a negative value and thus

$$\begin{aligned} \frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} &< 0 \\ \frac{p_{k-1}}{q_{k-1}} &< \frac{p_k}{q_k}, \end{aligned}$$

showing that every odd convergent is larger than its preceding even convergent. If we look at even k,

$$\begin{aligned} \frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} &> 0 \\ \frac{p_{k-1}}{q_{k-1}} &> \frac{p_k}{q_k}, \end{aligned}$$

showing that every odd convergent is larger than its proceeding even convergent.

To summarize, Corollary 4 tells us that every odd convergent has a value greater than both the preceding and proceeding even convergents. Corollary 6 tells us that the odd convergents form a decreasing sequence and the even convergents for an increasing sequence. By combining these two concepts, we end up with figure 1. This figure shows the odd and even convergent

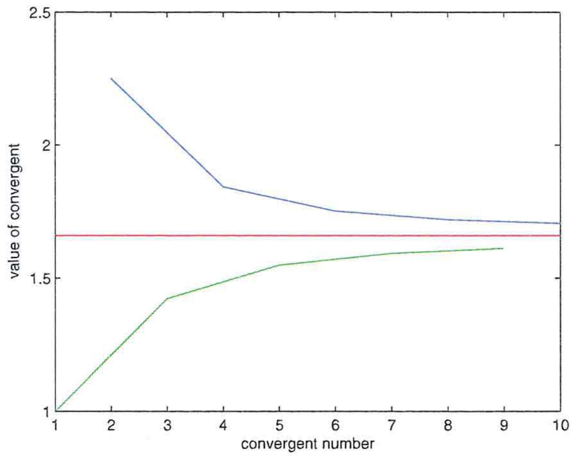


Figure 1: even and odd convergents tend to central line. Odd convergents = blue, Even convergents = green.

sequences tending toward a central value. This central value is the value of the continued fraction itself.

Another interesting point is that for $k \geq 2$ the sequence of q_k is strictly increasing:

Given equation (8) $q_k = a_k q_{k-1} + q_{k-2}$, we have $q_k > a_k q_{k-1}$ as q_{k-2} is positive (with smallest possible value $q_0 = 1$) and so $q_k > q_{k-1}$ as a_k is natural. So $\lim_{k\to\infty} q_k \to \infty$. This quality is necessary for the convergence of an infinite continued fraction, following Corollary 4, because then $\lim_{k\to\infty}\left(\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}}\right) = 0$, ie. $\lim \frac{p_k}{q_k} = \lim \frac{p_{k+1}}{q_{k+1}}$.

3 Convergence of the elements

When looking at infinite continued fractions convergence, I wondered if it was possible to form a non-convergent infinite continued fraction ie. one that does not have a defined value. If it is possible, what would happen to the sequences of the odd and even Convergents? Following Theorem 1 it seems likely that every infinite continued fraction with natural elements is defined, and is irrational. So I look for the non-convergent type of continued fraction with other positive, real elements.

Up until now, we have assumed that the elements take natural values, due to the process used for finding the simple continued fraction expansion of a number, shown in the first section. If instead of restricting the elements to taking only natural values, we allow them to take any positive values, the question of convergence of the sequence of convergents needs answering. Theorem 10 in [7] deals with this:

Theorem 7 ([7, Theorem 10]). *For the infinite continued fraction $[a_0, a_1, a_2, \ldots]$ to converge, it is necessary and sufficient that the series*

$$\sum_{n=0}^{\infty} a_n$$

should diverge.

Corollary 8. *An infinite continued fraction with natural elements will always represent a defined value and this value is an irrational number.*

Proof. Every infinite continued fraction with natural elements will have $\sum_{n=0}^{\infty} a_n \geq \sum_{k=0}^{\infty} 1$, as 1 is the smallest possible natural number. The series $\sum_{k=0}^{\infty} 1$ is divergent, meaning that all series $\sum_{n=0}^{\infty} a_n$, $a_n \in \mathbb{N}$, are divergent. Theorem 7 tells us that these continued fractions are convergent and by Theorem 1 they are irrational. □

Corollary 8 rules out natural element infinite continued fractions as possible non-convergent continued fractions.

According to Theorem 7 we need the elements to form a convergent series to get a non-convergent continued fraction. I am interested in what happens to the sequence of odd and even convergents of non-convergent continued fractions.

The geometric series is a simple series that I will use in order to show what happens.

Geometric series:

$$\sum_{k=0}^{\infty} mr^k = m + mr + mr^2 + \ldots + mr^k + \ldots \ (m \neq 0)$$

Converges if $|r| < 1$ and diverges if $|r| \leq 1$

If the sum converges,

$$\sum_{k=0}^{\infty} mr^k = \frac{m}{1-r}$$

[1, page 64]

For the purpose of this report, let $m = 1$ without loss of generality.

To form a continued fraction with geometric series elements, we let $a_k = r^k$, so $\sum_{k=0}^{\infty} a_k = \sum_{k=0}^{\infty} r^k$.

Due to the nature of the geometric series, the case $r < 1$ will be our main interest but cases $r \geq 1$ and $r = 1$ will also be investigated. Table 2 shows one of each case, and what is expected for the convergence of the corresponding continued fraction by Theorem 7.

Table 2

r	Series	C.frac by $Th^m 7$
0.5	convergent	non-convergent
2	divergent	convergent
1	divergent	convergent

Using MATLAB I created a function to make the continued fraction for each case of r in Table 2 and plotted the convergent sequences resulting ($\frac{p_k}{q_k}$).

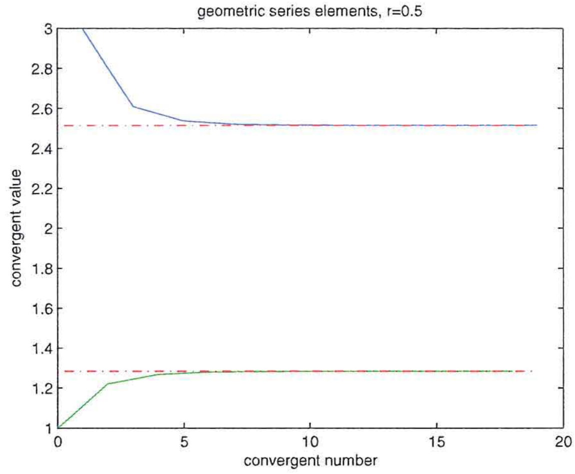


Figure 2: The "divergence" of the convergents for Geometric series elements, $r = 0.5$. Blue = odd convergents, Green = even convergents, Red dotted = limit of sequences.

Figure 2 shows the "divergence" of the convergents of this continued fraction. The odd convergents and even convergents each converge to their own separate values (2.5154 - odd, 1.2851 - even, $4dp$). Meaning that if we take the limit as $k \longrightarrow \infty$ for the convergent $\frac{p_k}{q_k}$ of this continued fraction, we find that it is undefined as the odd-even sequence is periodic.

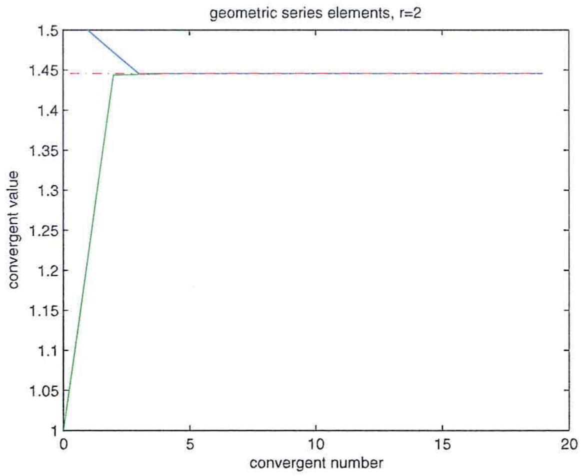


Figure 3: The convergence of the continued fraction for Geometric series elements, $r = 2$. Blue = odd convergents, Green = even convergents, Red dotted = limit of sequences.

Figure 3 shows the convergence of this continued fraction with $r = 2$ to a value of 1.4459 ($4dp$), the convergence is quite fast, reaching this value (to a $4dp$ accuracy) on convergent $\frac{p_3}{q_3}$. Note that this time, the sequences have the same limit, so this continued fraction **is** convergent.

Figure 4 shows the convergence of the continued fraction $[1; 1, 1, 1, \ldots]$. It seems to have a slower convergence than $r = 2$ case. In fact it has the slowest convergence of all continued

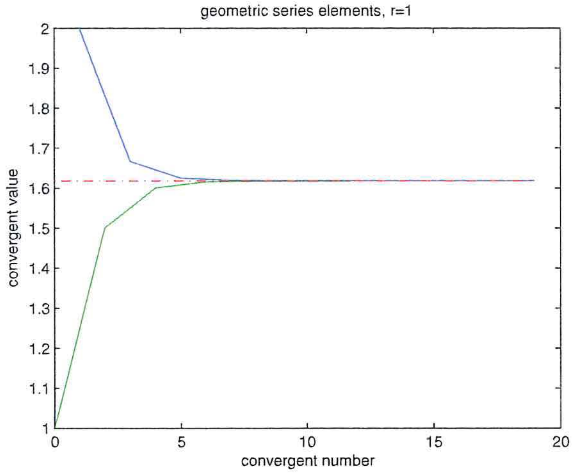


Figure 4: The convergence of the continued fraction for Geometric series elements, $r = 1$. Blue = odd convergents, Green = even convergents, Red dotted = limit of sequences.

fractions with divergent element geometric series. Again, both odd and even convergents have the same sequence limit. As we know from the first section, this continued fraction is the golden ratio and converges to $\Phi = 1.6180339\ldots$

The lack of convergence of case $r = 0.5$ can be explained using the following ideas:

1. The fact that the odd and even convergents sequences each form convergent sequences is explained by Corollary 6. Given $\frac{p_{k-2}}{q_{k-2}} - \frac{p_k}{q_k} = \frac{(-1)^{k-1}a_k}{(q_k q_{k-2})}$, we note that $a_k = r^k$ and $r < 1$ in this case. Thus as $k \to \infty$, $a_k \to 0$. So $\lim\limits_{k\to\infty} \frac{p_k}{q_k} = \lim\limits_{k\to\infty} \frac{p_{k-2}}{q_{k-2}}$. Thus the odd and even convergent sequences have convergence. They do not however converge to the same value.

2. As part of the proof for Theorem 7 in Khinchin's book [7], it is shown that if $\sum\limits_{k=0}^{\infty} a_k$ is a convergent series, then $q_k q_{k+1} < C$, where C is a finite, positive constant. Taking Corollary 4, this means that $\lim\limits_{k\to\infty} \left(\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k}\right) > \frac{(-1)^k}{C}$. Explaining the separation between the odd and even convergents sequences as $\lim\limits_{k\to\infty} \frac{p_k}{q_k} \neq \lim\limits_{k\to\infty} \frac{p_{k-1}}{q_{k-1}}$.

In general, for a series $\sum\limits_{k=0}^{\infty} a_k$ to be convergent, it is obvious that as $k \to \infty$, the k^{th} element must tend to 0. Each convergent sequence (odd and even) converges to a limit, with, as explained in the first section, the odd convergents forming a decreasing sequence and the even convergents forming an increasing sequence. Whether or not these limits are the same depends on the series of elements of the continued fraction in question, as stated in Theorem 7. If the limits are the same for odd and even convergent sequences, we have a convergent continued fraction with a defined value.

4 Invariant Density Function of the G-K-W Operator

Now we return to having elements $a_1, a_2, \ldots$ with natural values. We would like to know something about the distribution of these elements. It turns out that they exhibit a remarkable degree of statistical regularity: for a "typical" irrational number,

$$\mathrm{Prob}(a_i = w) = \frac{\ln \frac{(w+1)^2}{w\,(w+2)}}{\ln(2)}.$$

In this section I explain this incredible fact: we are led via a point transformation of the interval $[0,1]$ to the "Gauss-Kuzmin-Wirsing (G-K-W) operator" [5] (which maps functions on $[0,1]$ to functions on $[0,1]$) and its fixed point (the density of a "Gauss measure"). Along the way we need a bit of *ergodic theory*.

First, note that if $x = [0; a_1, a_2, a_3, \ldots]$ then the first element is easy to read off: it is $a_1 = \lfloor \frac{1}{x} \rfloor$. Since the first a is easily obtained, we iteratively apply the following procedure: delete the first a_i, shift the remaining ones along by one, then read off the (new) first element. A transformation T on $x \in [0,1]$ is defined by this process:

$$x = [0; a_1, a_2, a_3, \ldots] \mapsto [0; a_2, a_3, a_4, \ldots] := T(x).$$

So,

$$x = [0; a_1, a_2, a_3, \ldots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}} = \frac{1}{a_1 + \frac{1}{[0; a_2, a_3, \ldots]}} = \frac{1}{a_1 + T(x)}.$$

Then, $a_1 + T(x) = \frac{1}{x}$ so rearrangement yields $T(x) = \frac{1}{x} - a_1$. We are working with natural elements, so element a_1 is the largest integer not exceeding $\frac{1}{x}$, the transformation can be rewritten:

$$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \tag{16}$$

This is called the Gauss transformation.

The G-K-W operator's action on density functions is induced by T. One takes a density of points in $[0,1]$, and "pushes forward" each point by T. If the initial density function is f_0, the G-K-W operator gives a resulting density $[Gf_0]$ according to [7]:

$$[Gf_0](x) = \sum_{k=1}^{\infty} \frac{1}{(x+k)^2} f_0\left(\frac{1}{x+k}\right) \tag{17}$$

We can define the initial density function on the interval [0,1], f_0 by:

$$\int_{x_0}^{x_0+\delta x} f_0(x)dx = \text{the proportion of numbers in } [x_0, x_0+\delta x] \text{ relative to } [0,1],$$

or equivalently, it is the probability of finding a randomly chosen number from the interval [0,1] in $[x_0, x_0 + \delta x]$. As we take $\delta x \to 0$, $f_0(x)$ tends to its true initial continuous probability density. This initial density is obviously $f_0(x) = 1$, because the numbers on [0,1] are distributed uniformly. The initial density is defined to be a probability density, thus iteration of the G-K-W operator will also yield a probability density as the total number of points is not lost by this process.

For most operators, there is a density that the operated initial densities, $G^i f_0$, tend towards as $i \to \infty$. This density is known as an invariant density as it is not changed by further application of the operator. It was realized by Gauss that the invariant density of this operator is given by [5][7, page 81]:

$$f_* = \frac{1}{\ln(2)} \frac{1}{(1+x)} \tag{18}$$

Gauss' proof was never published, whether because he never actually proved this, we do not know. The first published proof came from Kuzmin in 1928 [7]. Kuzmin's analytical proof that gives (18) is rather long and complicated (involving measure theory), so will not be looked at here.

Gauss was also interested in the relationship between the invariant density and operated initial density, for large n (n =number of iterations of operator G), but was not known to have come up with a solution [7, page 81]. Kuzmin found a solution, showing that:

$$G^n f_0 = f_* + e(n) \tag{19}$$

where the difference $e(n) = G^n f_0 - f_*$, has $||e(n)||$ bounded by $O(e^{-S\sqrt{n}})$ for some $S > 0$ [7](page 83).

It has since been found that the difference typically has $||e(n)|| = ||G^n f_0 - f_*|| \leq K(e^{-Sn})$ for some K dependent on the initial density and so $||e(n)|| = O(e^{-Sn})$ [5][4]. I wish to illustrate this latter statement.

Instead of delving through complicated theory, I approximate the operator numerically using MATLAB in order to find the invariant density. I then attempt to show numerically that the rate at which an initial distribution transforms into the invariant density is $O(e^{-Sn})$. I also look into the dynamical "orbit" given by the Gauss transformation on points in [0,1].

4.1 The invariant density - First attempt

In my initial attempt to find the invariant density numerically, I split the interval [0,1] into N uniform, subinterval "*bins*". Let the i^{th} *bin* be denoted by B_i. I generated p random numbers (points) on the interval [0,1] and I had MATLAB arrange these numbers into the N *bins* of size $\frac{1}{N}$, arranged by the value of x: $x \in B_i : \frac{i-1}{N} < x \leq \frac{i}{N}$, $i = 1, \ldots, N$. This gives a "pixelated" approximation to the initial distribution. To make an initial *probability* density, I normalized the distribution by dividing the number of "points" in each *bin* by p, thus finding the fraction of the total points in each *bin* and I plotted the resultant density in a histogram. This histogram gave an approximately constant initial density f_0, with slight differences due to random number errors. I then had MATLAB apply the Gauss transformation to each randomly chosen point and re*bin* the resulting numbers, dividing by p again to get their probabilities. I repeated this process for a large number of iterations compared to the number of *bins* used. This is in effect, an approximation to the result, of iterated application of the G-K-W operator to a uniform initial density function.

Results

The results had a certain amount of error due to the limit of MATLAB's accuracy with the random numbers chosen. This error from the theorized invariant (probability) density is seen in figure 8 as a *noise* centered about the theorized density after 1000 iterations of my process described above.

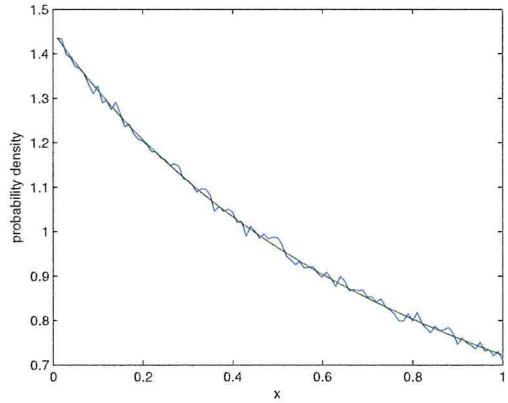


Figure 5: The numerically found invariant density, with p = 1000000, N = 100, n = 1000 iterations, plotted with the theorized invariant density. Green = theoretical invariant density, blue = numerically found invariant density found.

Plotting a graph of the absolute sum of differences between the theorized density and the transformed initial density in each *bin* versus iteration number results with figure 6.

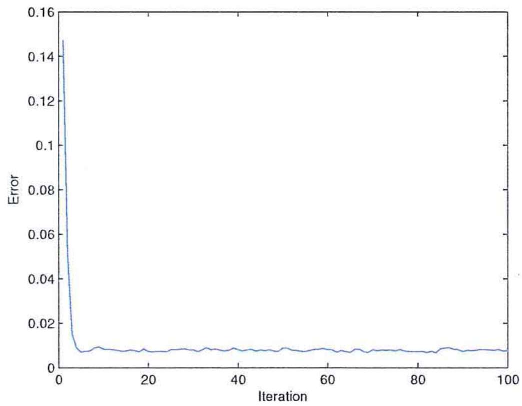


Figure 6: The error versus iteration number for attempt 1

Figure 6 shows a rapid decrease down to an above zero error. This "final" error is greater than zero due to the finite number of points used, causing the statistical "noise". Again, this "noise" is due to MATLAB's accuracy with this method and the number of points used. This first attempt does not give much useful data due to the accumulation of round off error after multiple transformations.

4.2 The invariant density - Second attempt

In an attempt to get better results, I reduced the effect of roundoff error by making a change to the numerical method used to find the invariant density function. To do this I use an N by N matrix P to approximate the linear G-K-W operator numerically. I still split the interval [0,1] up into N subintervals as before. I note that the G-K-W operator basically shifts the "area" under the density curve to form a new density each iteration, so we can look at probabilities associated with moving "area" from one subinterval to another. P_{ij} holds the probability of the Gauss transformation taking a number from B_i to B_j, where B_i is defined as it was in the first attempt (the i^{th} *bin*). This matrix method is known as Ulam's method [6][9]. MATLAB was used once again, but this time applying the stochastic matrix P to the initial probability density.

Creating this matrix with precise entries is complicated, involving the knowledge of many elements of the continued fraction expansion of the numbers in [0,1]. Instead I approximated these probabilities in the following way. I generated m random numbers for each of the N *bins*. The Gauss transformation was then applied to the numbers in each of these *bins*, and the transformed numbers redistributed into their new *bins*. The fraction of numbers in B_i moved to B_j was calculated, and placed at this position in the matrix at P_{ij}. This process only involves one iteration of the Gauss transformation so has a reduced roundoff error in comparison with attempt 1. Note that increasing m would make the probability entries of P more precise by the Central Limit Theorem and this should be seen in the results.

Theory of matrix P

The eigenvalues and corresponding eigenvectors of P are of great interest in the search for the invariant density function. P is stochastic and because of this, the Perron-Frobenius Theorem [13] asserts that the dominant eigenvalue of P is $\lambda_1 = 1$ and that all other eigenvalues have $|\lambda_i| < 1$, for $i = 2, 3, \ldots, N$. The theorem also asserts that along with λ_1, there is a corresponding positive eigenvector whose sum is 1. This eigenvector is a scalar multiple of the approximate invariant density, denoted by v_1. The approximate invariant *probability* density is given by Nv_1 having been normalized so that the area beneath Nv_1 on [0,1] is 1. To show that Nv_1 is the approximate invariant probability density, we know that $\lambda_1 = 1$ is Nv_1s eigenvalue as Nv_1 is in the same eigenspace as v_1 and so we have $PNv_1 = \lambda_1 Nv_1 = 1Nv_1 = Nv_1$. Thus Nv_1, a scalar multiple of the dominant eigenvector, is the invariant probability density of the matrix approximation to the operator, as it is not changed by application of the G-W-K operator approximation, P.

Using MATLAB I compared the approximate invariant probability density Nv_1 with the theorized invariant probability density f_* and found that the error between them is reduced with increase in the value of m (number of points per *bin*), but reduced further when N (number of *bins*) is also increased. This is expected, since as $N \to \infty$ the finite dimensional eigenvector Nv_1, will tend towards an infinite dimensional vector, a function. As $m \to \infty$, the probabilities of the matrix P tend to their correct values, as we remove more statistical inaccuracies by involving the entire population of numbers on [0,1]. The result of increased N giving lower error is expected by Murray's paper [9], it is said that if we denote by h_N the invariant probability density given by the N by N approximation matrix (my Nv_1), then

$\lim_{N\to\infty} h_N = h$, where h is the true invariant probability density of the G-K-W operator. So, as we increase N we in a rough sense, have P becoming more like the G-K-W operator.

P has $N-1$ other eigenvalues and $N-1$ corresponding other eigenvectors, most involving complex numbers. If the "spectrum" of the eigenvalues is plotted in the complex plane, one finds a dense circle of eigenvalues centered on $0+0i$ with the second absolute largest eigenvalue $\lambda_2 =$ -0.303.... Figure 7 shows this spectrum for $N = 100$, and initially 1000000 points per *bin*. The eigenvalue λ_2, defined as the second largest eigenvalue of P, has $|\lambda_2|$ equal to the value of the "non peripheral spectral radius". This result was also found by Sebe[12] and coincides with the *Wirsing* constant $\lambda = 0.30366\ldots$. The Wirsing constant is the second largest eigenvalue of the G-K-W operator[5].

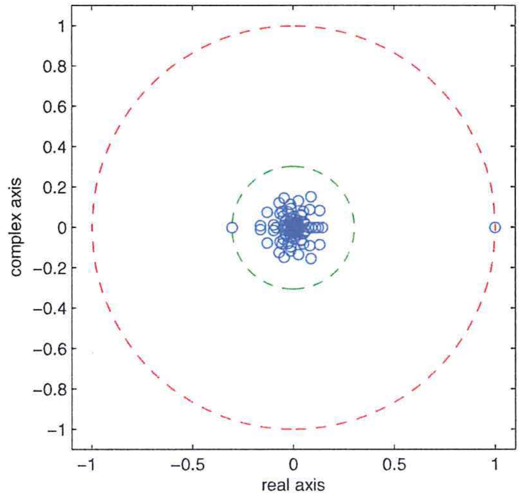


Figure 7: The spectrum of eigenvalues for $N = 100$, $n = 1000000$. Blue = individual eigenvalues, Red = circle of radius $|\lambda_1|$, Green = circle of radius $|\lambda_2|$.

As with attempt 1, the difference between the invariant density function and the initial density function after the n^{th} iteration of the operator, can be thought of as an error of sorts. Remember I am trying to show that this error is $O(e^{-Sn})$, $S > 0$, numerically.

Let the initial probability density f_0 have the finite N dimensional vector approximation f_0'. Given the initial probability density approximation f_0', we can rewrite f_0' as a linear combination of the eigenvectors of P following Poole [11, Theorem 4.19]. Thus

$$f_0' = c_1 v_1 + c_2 v_2 + c_3 v_3 + \ldots + c_N v_N$$

So that:

$$\begin{aligned}
P f_0' &= P(c_1 v_1 + c_2 v_2 + c_3 v_3 + \ldots + c_N v_N) \\
&= P c_1 v_1 + P c_2 v_2 + P c_3 v_3 + \ldots + P c_N v_N) \\
&= c_1(P v_1) + c_2(P v_2) + c_3(P v_3) + \ldots + c_N(P v_N) \\
&= c_1(\lambda_1 v_1) + c_2(\lambda_2 v_2) + c_3(\lambda_3 v_3) + \ldots + c_N(\lambda_N v_N) \,.
\end{aligned}$$

If we iterate this process n times:

$$P^n f_0' = c_1(\lambda_1^n v_1) + c_2(\lambda_2^n v_2) + c_3(\lambda_3^n v_3) + \ldots + c_N(\lambda_N^n v_N) \,. \tag{20}$$

It happens that for all initial probability densities, $c_1 = N$ (from equation (20)). This can be shown as follows: Because P is stochastic, it is "Markov", ie. if $\|f_0'\|_1 = N$ then $\|Pf_0'\|_1 = N$ and $\|P^n f_0'\|_1 = N$ for any $n = 1, 2, \ldots$. Where the one-norm $\|.\|_1$ is defined as $\|u\|_1 = |u_1| + |u_2| + |u_3| + \ldots + |u_k|$ for $u = [u_1, u_2, u_3, \ldots, u_k]^T$. (When we find $\|f_0'\|_1$, we are in effect summing the heights of the initial probability density f_0' across all subintervals on [0,1], so the total height is N because this gives the subintervals an average of 1 as there are N subintervals, thus giving a total probability of 1 as expected).

Following equation (20)

$$P^n f_0' = c_1(\lambda_1^n v_1) + \sum_{i=2}^{N} c_i(\lambda_i^n v_i) \; .$$

So,

$$\|P^n f_0'\|_1 = \left\| c_1(\lambda_1^n v_1) + \sum_{i=2}^{N} c_i(\lambda_i^n v_i) \right\|_1 \; .$$

Which as $n \to \infty$ reduces the terms $\sum_{i=2}^{N} c_i(\lambda_i^n v_i)$ to zero vectors and gives

$$\lim_{n \to \infty} \|P^n f_0'\|_1 = |c_1| \|v_1\|_1 \tag{21}$$

following the norm axioms and because $\lambda_1 = 1$ and $|\lambda_i| < 1$ for all $i = 1, 2, \ldots, N$. Given my statement that $\|P^n f_0\|_1 = N$ and also because $\|v_1\|_1 = 1$ (as a result of the Perron-Frobenius Theorem), $|c_1| = N$. Not only are the one-norms of these vectors equal to N, but the sums are as well because we cannot have negative probability densities, thus $c_1 = N$.

Having $c_1 = N$ for **any** initial probability density, means that no matter what probability density we start with, application of the approximate G-K-W operator to this density, in the limit tends to the approximate invariant density function, by equation (20) as $\lim_{n \to \infty} P^n f_0' = c_1 v_1 = N v_1$. As N, the number of *bins*, tends to infinity, this approximate density approaches the true density $f_* = \frac{1}{\ln(2)} \frac{1}{(1+x)}$ [9].

Exponential Decay of Error

Given that $c_1 = N$ and $\lambda_1 = 1$, $P^n f_0'$ can be written $P^n f_0' = N v_1 + \sum_{i=2}^{N} c_i(\lambda_i^n v_i)$. Because $P^n f_0'$ is my approximation of $G^n f_0$ and $N v_1$ my approximation of f_*, the difference between the n^{th} operated initial probability density and the invariant probability density can be approximated by $e(n)' = P^n f_0' - N v_1 = \sum_{i=2}^{N} c_i(\lambda_i^n v_i)$.

Now that some theory has been established I can look at the effect iteration of P has on the error. For this numerical method I have taken the error to be the one-norm of this difference,

$\|e(n)'\|_1 = \|P^n f_0' - Nv_1\|_1$.

$$\begin{aligned}
\|e(n)\|_1 &= \left\| \sum_{i=2}^{N} c_i(\lambda_i^n v_i) \right\|_1 \\
&\leq \|c_2 \lambda_2^n v_2\|_1 + \sum_{i=3}^{N} \|c_i(\lambda_i^n v_i)\|_1 \\
&= |c_2 \lambda_2^n| \|v_2\|_1 + \sum_{i=3}^{N} |c_i \lambda_i^n| \|v_i\|_1
\end{aligned}$$

(Following the norm axioms).

The decay of $\|e(n)'\|_1$ is dominated by the term $|c_2\lambda_2^n|\|v_2\|_1$ as $|\lambda_2| > |\lambda_i|$. It follows that:

$$\begin{aligned}
\|e(n)'\|_1 &\leq \left(|c_2| \|v_2\|_1 + \sum_{i=3}^{N} \left| c_i \left(\frac{\lambda_i}{\lambda_2} \right)^n \right| \|v_i\|_1 \right) |\lambda_2^n| \\
&\leq \left(|c_2| \|v_2\|_1 + (N-2) \max_{i=3\ldots N} \left\| c_i \left(\frac{\lambda_i}{\lambda_2} \right)^n v_i \right\|_1 \right) |\lambda_2^n| \\
&\leq \left(|c_2| \|v_2\|_1 + (N-2) \max_{i=3\ldots N} \|c_i v_i\|_1 \right) |\lambda_2^n| \,.
\end{aligned}$$

because $\left|\frac{\lambda_i}{\lambda_2}\right| \leq 1$.

So we can define a positive constant $K \leq \left(|c_2| \|v_2\|_1 + (N-2) \max_{i=3\ldots N} \|c_i v_i\|_1 \right)$ such that:

$$\|e(n)'\|_1 = K|\lambda_2^n|$$

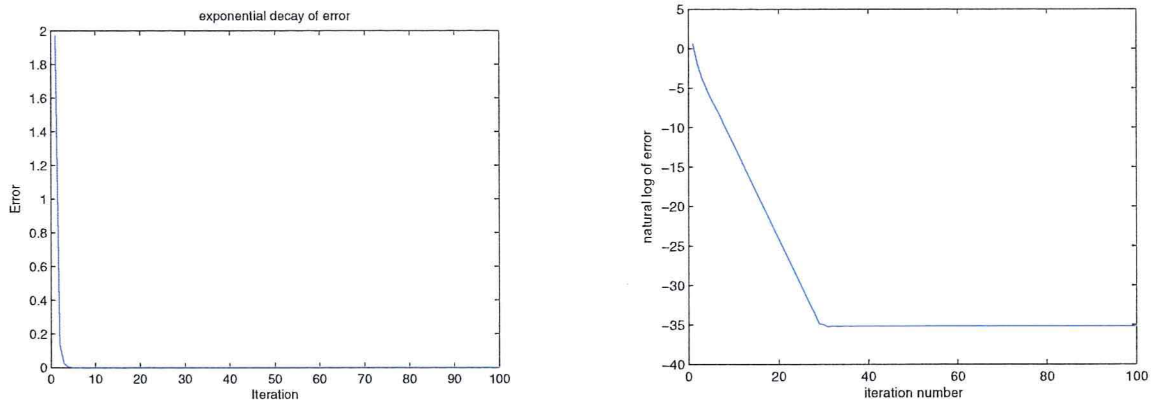
Experimental Results

I used MATLAB to work through n iterations of the approximation of the G-K-W operator on the initial probability density approximation f_0' and I plotted the results of the error versus iteration number. The slope appeared to be exponential, as conjectured. To check, I then plotted the natural log of the error versus iteration number and found an approximate straight line, finding the gradient and y-intercept by least squares. Figures 8a and 8b display this.

Note: at about $n = 30$ Figure (8b)'s plot flattens out to a horizontal line, this is because it has reached MATLAB's accuracy. If MATLAB was not limited, the line would continue at a constant gradient for all n, by equation (23).

$$\begin{aligned}
\|e(n)'\|_1 &= K|\lambda_2^n| && (22) \\
\ln(\|e(n)'\|_1) &= \ln(|\lambda_2|)n + \ln(K) && (23)
\end{aligned}$$

From MATLAB's least squares output values, I found the gradient of $\ln(\|e(n)'\|_1)$ versus n to be very close to $\ln|\lambda_2|$ and if we neglect the first n_0 terms, we find that least squares gives a



(a) Exponential decay of error, $N = 100$,

(b) Natural log of error

Figure 8: The decay of error with iteration n

slope value equal to $\ln|\lambda_2|$. This is because by iteration n_0, the terms involving the other eigenvalues have reduced and become negligible. If we now take the natural exponential of equation (23) we have:

$$\|e(n)'\|_1 = Ke^{\ln(|\lambda_2|)n} \tag{24}$$

So I have illustrated numerically that the error between the invariant probability density and the probability density after n applications of the G-K-W operator, decays $\mathrm{O}(e^{(-Sn)})$ where $S = -\ln(|\lambda_2|)$. This is the rate at which any initial density transforms into the invariant density. An example of a value found for λ_2 is $\lambda_2 = -0.3039$ for $N = 100$, $p = 1000000$, which is the value of Wirsing constant (-0.3036...) to 3 significant figures. Remember, the Wirsing constant is the second absolute largest eigenvalue of the G-K-W operator, as said earlier.

I found that increasing m, N or both gives a value of S such that λ_2 becomes closer to the value of the Wirsing constant. From this result it seems reasonable that the true operator would have an analogous decomposition of the initial density function f_0 into a linear combination of eigenfunctions and that the error would be dominated by the term involving the Wirsing constant, giving the same result of decay, $O(e^{-Sn})$.

4.3 Dynamics of Continued Fractions under the Gauss Transformation

By Birkhoff's Ergodic Theorem [2, page 96], if we take a "typical" randomly chosen point $x \in [0, 1]$ and transform it using the Gauss transformation many times, the density of all the transformed points will also follow the invariant density distribution $\frac{1}{\ln(2)}\frac{1}{(1+x)}$ (where "typical" assumes the point taken is irrational, and non-periodic). This concept is illustrated in this subsection.

I look at the mappings $x, T(x), T^2(x), \ldots, T^n(x), \ldots$, known as the "orbit" or "trajectory" of the transformed points [2, page 20]. These orbits can be looked at in what is know as a **cobweb** diagram. A cobweb diagram is made using the following process:

1. Draw the transformation curve $T(x) = \frac{1}{x} - \lfloor\frac{1}{x}\rfloor$. $\rightarrow$ 2. For starting point x_0 begin at (x_0, x_0) on the graph. $\rightarrow$ 3. Draw a line vertically to $(x_0, T(x_0))$. $\rightarrow$ 4. draw a line horizontally to $(T(x_0), T(x_0))$. $\rightarrow$ 5. Repeat from 2. with starting point $x_1 = T(x_0)$, etc.

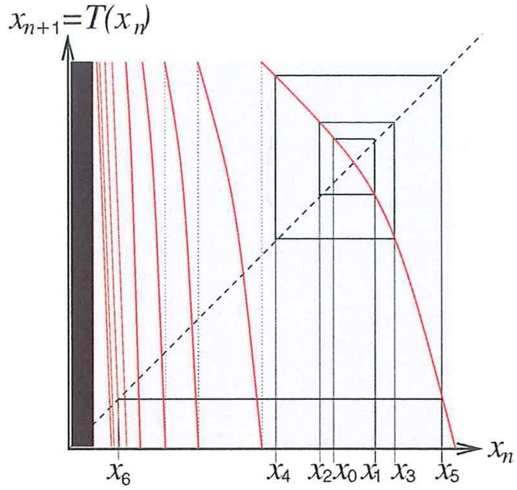


Figure 9: A cobweb diagram example taking x_n to $x_{n+1} = T(x_n)$

The cobweb process takes the point x_0 on the x-axis to its transformed point $x_1 = T(x_0)$ back onto the x-axis and when continued with further transformations, $x_2 = T^2(x_0), x_3 = T^3(x_0), \ldots,$ leaving a "cobweb" showing the transformed points. In general, a randomly chosen value x_0 on the interval [0,1] leads to a chaotic cobweb diagram similar to that in figure 9. This means that there will be no repeating parts in the orbit, it is completely erratic.

To investigate what happens to the orbit in the long run, we can subdivide the interval [0,1] into N subintervals of size $\frac{1}{N}$, as in the previous methods and again denote the i^{th} subinterval by B_i. The relative frequency of these mappings landing in a given subinterval B_i is given by [2, page 6] [8, page 3]:

$$t_i(x_0, n) := \frac{1}{n(\text{interval size} = \frac{1}{N})}\{\text{number of } T^j(x_0) \in B_i, j = 1, \ldots, n\}. \tag{25}$$

If we take the limit as $n \rightarrow \infty$, $t_i(x_0, n)$ tends to what is known as the "time" averaged probability density of being in the subinterval B_i. Birkhoff's ergodic theorem tells us that this time averaged density for "typical" points on [0,1], should also be the invariant density of the G-K-W operator. MATLAB is used to illustrate that this is the case.

Using MATLAB I tracked a random non-periodic irrational number's orbit for n iterations and plotted the resultant density using equation (25) in a histogram with the N uniform subintervals B_i. The result was approximately the invariant density of the G-K-W operator, as seen in figures 10, 11a and 11b. Figure 10 shows a slight amount of variation between the invariant density and the numerical histogram created. Figure 11a shows the effect of increasing N by itself, although more subintervals, we have the same number of points filling them, giving more error due to the lower ratio of $\frac{n}{N}$. Figure 11b shows the result of increasing both N and n, resulting in a histogram with a much closer fit to the invariant density. We cannot achieve 100% precision with MATLAB however because MATLAB can only achieve a low ratio of total points to number of histogram bins when using a high number of bins, due to limited memory. But by extrapolation of these results it is expected that as both N and $\frac{n}{N}$ $\to \infty$ we would result in the invariant density.

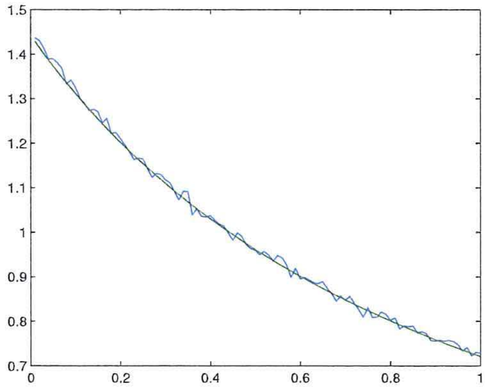
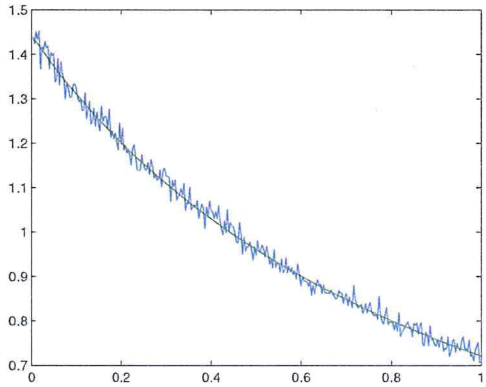
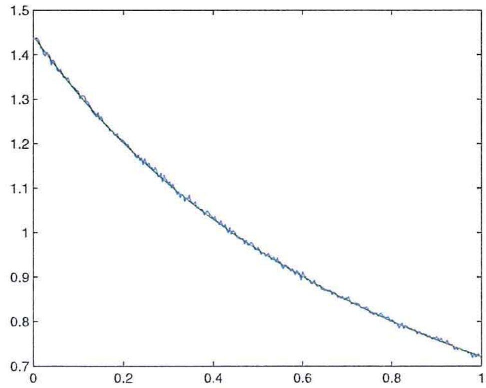


Figure 10: The time averaged orbital density for a single random number, iterations $n = 10^6$, number of *bins* $N = 100$.



(a) $n = 10^6$, $N = 300$.



(b) $n = 10^8$, $N = 300$.

Figure 11: The time averaged orbital density for a single random number, varied parameters.

The continued fraction expansion of any x in [0,1], has first element $a_1 = w$ if $\frac{1}{w+1} < x \leq \frac{1}{w}$. Thus generalizing, a continued fraction has any element $a_i = w$, $i = 1, 2, 3, \ldots$, after $i-1$ iterations of the Gauss transformation if $\frac{1}{w+1} < T^{i-1}(x) \leq \frac{1}{w}$. Because the time averaged probability density of x's orbit is the invariant density in the case of a "typical" number, we can find the probability that *any* element in a "typical" continued fraction expansion is w. It follows that probability that $a_i = w$ (for any $i = 1, 2, \ldots$) is given by:

$$\text{Prob}(a_i = w) = \int_{\frac{1}{1+w}}^{\frac{1}{w}} \frac{1}{\ln(2)} \frac{1}{(1+x)} dx \tag{26}$$

$$= \frac{\ln(1+\frac{1}{w}) - \ln(1+\frac{1}{w+1})}{\ln(2)} \tag{27}$$

$$= \frac{\ln \frac{(w+1)^2}{w(w+2)}}{\ln(2)} \tag{28}$$

The probabilities for the first 7 values for any element are given in table 3:

Table 3

number w	1	2	3	4	5	6	7
probability	0.4150	0.1699	0.0931	0.0589	0.0406	0.0297	0.0227

Table 3 shows that the element value $w = 1$ is visited most often in the orbit of a "typical" randomly chosen continued fraction in [0,1], following the Gauss transformation, and that the sequence $w = 1, 2, 3, \ldots$ has a decreasing probability. Although the probability density is greater toward zero, and thus there is a greater chance of finding $T^i(x)$ for any $i = 1, 2, \ldots$ closer to zero per *uniform* subinterval, the intervals involving the elements are *non-uniform*, such that the most likely value for any a_i is 1.

Although we get the invariant density for most ("typical") initial x, by Birkhoff's Ergodic Theorem [2, page 96], some continued fractions (a minority) do not yield the invariant density as their time averaged orbital density. An example is periodic continued fractions.

Periodic Continued Fractions

Periodic continued fractions are continued fractions whose elements after the k^{th} element (for some integer $k \geq 0$) repeat periodically.

$$\alpha_{periodic} = [a_0; a_1, a_2, \ldots, a_k, \overline{a_{k+1}, a_{k+2}, \ldots, a_{k+l}}] \tag{29}$$

To begin with, I am only interested in the effect of the Gauss transformation $T(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$, on the periodic part of $\alpha_{periodic}$ so we will only look at periodic continued fractions on the interval [0,1] lacking the transient non-periodic elements. In the end, I will generalize my findings to include the periodic continued fractions with transient elements.

$$x_{periodic} = [0; \overline{a_1, a_2, \ldots, a_m}] \tag{30}$$

We can find all the periodic continued fractions of this kind, for any natural value of m. Periodic continued fractions with one repeated element are mapped to themselves after one Gauss transformation. Periodic continued fractions with two elements repeated in order are mapped to themselves after applying the transformation twice. In general, a periodic continued fraction of the type described by (30) with m repeated elements are mapped to themselves after m successive applications of the transformation, ie. $T^m(x_0) = x_0$. A cobweb diagram of a periodic continued fraction of the type in (30) following the Gauss transformation takes closed path orbits, as the elements are repeated, unlike non-periodic irrationals.

Periodic continued fractions do not follow the invariant density as they have a periodic nature and will **not** visit very many different points on the interval, which is obviously necessary to give this density. We instead end up with m discrete points on the interval [0,1] corresponding to the m repeated elements of equation (30). Figure 12 displays this, with number of histogram bins, $N = 100$.

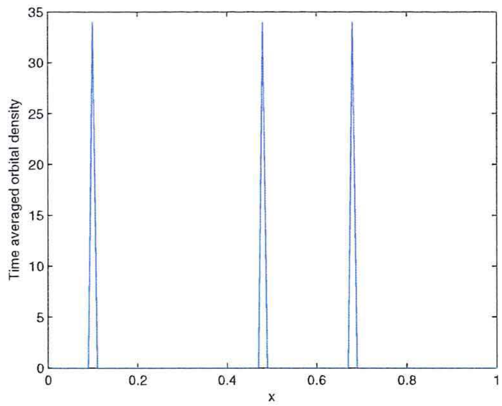


Figure 12: The time averaged orbital density of periodic continued fraction $x = [0, \overline{1, 2, 10}]$ following the Gauss transformation, $N = 100$.

If we now look at the orbit of the continued fractions of type (29), the result is a finite number of transient points, which then settle down to a periodic orbit. We still end up with a time averaged orbital density of the type in figure 12. This is because as $n \to \infty$, the transients have died away and become negligible, while the periodic parts take the majority of the orbital path. We can therefore generalize and say that all periodic continued fractions are excluded from having a time averaged probability density of $\frac{1}{\ln(2)} \frac{1}{(1+x)}$ and are "atypical" by this nature.

4.4 Interpretation of Results

Not only does the probability density $f_* = \frac{1}{\ln(2)} \frac{1}{(1+x)}$ represent the invariant density function of the G-K-W operator, it is also the "time" averaged probability density of the orbit of points following the Gauss transformation for "typical" points on [0,1].

This result gives a remarkable degree of statistical regularity for values of a "typical" continued fraction's elements. It means that every "typical" point in [0,1] has a continued fraction expansion with the same ratios of element values, with 1 occurring 41.50% of the time, 2 occurring 16.99% of the time, etc. (see Table 3). This suggests that all "typical" continued fractions under iterations of the Gauss transformation seem to "forget" their initial states. Whereas, periodic continued fractions do not "forget" their initial states and their time averaged orbital densities show their periodic nature.

If we combine all of the time averaged probability densities for orbits of all single points on [0,1] into an overall averaged density, we get the invariant probability density defined by the G-K-W operator. Given that randomly chosen points with periodic continued fraction expansions do not follow the invariant density, we accept that an overwhelming majority of the numbers on [0,1] must be non-periodic irrationals, as the periodic densities have a non-existent effect on the invariant density (periodic continued fractions have measure 0).

Out of the chaos of a "typical" orbit, comes an order in its probabilities.

5 Conclusion

This report has given an introduction to some basic concepts of continued fractions. It has given an example showing that for an infinite continued fraction with elements that form a convergent series, the continued fraction diverges, ultimately because the odd and even convergents converge to their own limits. I have illustrated via a numerical approximation, that the G-K-W operator has an invariant probability density given by $f_* = \frac{1}{\ln(2)}\frac{1}{(1+x)}$ and any initial probability density will tend to it with successive operations of this operator. The invariant density is also the "time averaged" orbital path of a non-periodic irrational continued fraction expansions and in consequence, periodic continued fractions form a minority of the numbers on the interval [0,1]. As a consequence, I found the probability of any element $a_i = w$, for a non-periodic continued fraction is given by $Prob(a_i = w) = \frac{\ln \frac{(w+1)^2}{w(w+2)}}{\ln(2)}$. Finally, I showed numerically that the decay of error between the invariant density and the n^{th} operated probability density for any initial probability density is $O(e^{-Sn})$, where $S = \ln|\lambda_2|$.

References

[1] H. Anton, I. Bivens, and S. Davis. *Calculus - Early Transcendentals.* John Wiley and Sons, inc., Eighth edition, 2005.

[2] A. Berger. *Chaos and Chance - An introduction to stochastic aspects of dynamics.* Walter de Gruyter GmbH & Co, 2001.

[3] C. Brezinski. *History of Continued Fractions and Pade Approximants.* Springer-Verlag, 1991.

[4] R. Burden and J. Faires. *Numerical analysisS.* Prindle, Weber and Schmidt, Third edition, 1985.

[5] P. Flajolet and B. Vallsee. Continued fraction algorithms, functional operators, and structure constants. Technical report, Institut National de Recherche en Informatique et Automatique, 1996.

[6] G. Froyland. Ulams method for random interval maps. *Nonlinearity*, 12:10291052, 1999.

[7] A. Khinchine. *Continued Fractions.* P. Noordhoff Ltd. Groningen, third edition, 1963.

[8] A. Lasota and M. C. Mackey. *Chaos, Fractals and Noise - Stochastic Aspects of Dynamics.* Springer-Verlag, Second edition, 1994.

[9] R. Murray. Approximation error for invariant density calculations. *Discrete and Continuous Dynamical Systems*, 14(3), July 1998.

[10] C. D. Olds. *Continued Fractions.* Random House Inc., 1963.

[11] D. Poole. *Linear Algebra: A Modern Introduction.* Thompson Brooks/Cole, Second edition, 2006.

[12] G. I. Sebe. On wirsings approach to gauss problem and related questions. *Politehn. Univ. Bucharest Sci. Bull. Ser. A Appl. Math. Phys.*, 1998.

[13] E. Seneta. *Non-negative Matrices and Markov Chains.* Springer-Verlag New York Inc, 1981.